The Annals of Statistics
1993, Vol. 21, No. 4, 2022–2042

ADAPTIVE ESTIMATION IN NONCAUSAL STATIONARY AR PROCESSES

BY E. GASSIAT

Université Paris-Sud

We consider the estimation problem of the parameter b of a stationary AR(p) process without any of the usual causality assumptions. The aim of the paper is to derive asymptotic minimax bounds for estimators of b. When the distribution of the noise is known, we show LAN properties of the model and derive locally asymptotically minimax (LAM) estimators. The most important results are about the case of unknown distribution: The main result shows that, if one uses the usual parametrization, these bounds depend heavily on the causality or the noncausality of the process, so that adaptive efficient estimation is impossible in the noncausal situation: The scaling factor is shown to give the hardest one-dimensional subproblem, and an unusual scaling is exhibited that could lead to adaptive efficient estimation of the rescaled parameter even in the noncausal case.

1. Introduction. In this paper we study optimal estimation for possibly noncausal autoregressive stationary processes. Here, $X = (X_t,\ t \in \mathcal{Z})$ satisfies the following autoregressive equation:

$$X_t + \sum_{k=1}^{p} b_k X_{t-k} = U_t, \qquad (1)$$

for all integers t.

We shall assume that p is the true order of the model, that is, $b_p \neq 0$. The U_t, $t \in Z$, form a sequence of independent random variables with common distribution F having finite variance. We define the polynomial B on the complex domain by

$$B(z) = 1 + \sum_{k=1}^{p} b_k z^k$$

and we make the following assumptions:

(A1) The polynomial B has no root with magnitude equal to 1 in the complex domain.
(A2) The distribution F is not Gaussian and U is centered.

(A1) is the assumption that ensures that stationary solutions of (1) exist and are uniquely defined.

Received June 1991; revised October 1992.
AMS 1991 *subject classifications*. Primary 62F35; secondary 86A15, 93E12.
Key words and phrases. AR processes, noncausal filtering, minimax estimation, adaptive estimators.

Under assumption (A2), given the processes X and $U = (U_t)$, $b = (b_k)$, $k = 1, \dots, p$, is uniquely defined by (1). The process X is then an autoregressive process of order p, for which the process U is not necessarily the innovation process. In other words, X_t and the U_{t+k}, $k > 0$, may be dependent. The process X is said to be causal (or minimum phase) if B has no roots inside the unit circle, since in this case X_t depends only on the present and the past of the noise process, $U_t, U_{t-1}, \dots$. X is said to be noncausal if it is not causal. More specifically, X is said to be purely noncausal if B has all its roots inside the unit circle; in this case, X_t depends only on the future of the noise process, $U_{t+1}, U_{t+2}, \dots$. Finally, if B has roots both inside and outside the unit circle, X is said to be mixed, and depends both on the past and the future of the noise process.

The problem we are interested in is the efficient estimation of the parameter b given observations $X_1, \dots, X_n$. We will study this problem when the distribution F of the driving noise U is known (in that case the problem is purely parametric) and also in the semiparametric case where F is unknown. On the subject of estimation of the parameter of an AR causal process, that is, for which U is the innovation process, there exists a vast literature. The noncausal case has rarely been studied, though applications exist in several domains such as telecommunications and geophysics; see Benveniste, Goursat and Ruget (1980) and Robinson (1984). Notice here that the nonstationary process defined by (1) and starting from x_0 is asymptotically the same as the stationary one if and only if b is causal. In the noncausal case, the nonstationary process is exploding, and b may be estimated using least-squares estimators; see, for example, Touati (1989). Notice also that any estimation method based on the second-order properties of the system will be unable to distinguish among causal and noncausal models, since the autocovariance function remains unchanged when the poles of $1/B$ are moved outside the unit circle.

In a previous paper we proposed a family of estimators and studied their asymptotic performances; see Gassiat (1990) and the references therein. These estimators were specifically based on higher moment structures of the system, for example, on the fourth cumulant. In the same manner, estimators based on the use of higher-order cumulant spectra were developed by Lii and Rosenblatt (1982). More recently, Breidt, Davis, Lii and Rosenblatt (1991) have studied the maximum likelihood estimator of b and the variance of U_t for possibly noncausal processes.

We are here interested in optimality criterion in the general noncausal autoregressive situation. In the causal case, Kreiss (1987a, b) showed the existence of adaptive estimators that achieve efficiency and gave the construction of such sequences. For this purpose, he used the concept of local asymptotic normality (LAN) of Le Cam (1960) and the local asymptotic minimax (LAM) theory developed for such models, introduced by Hájek (1972) and exploited by Fabian and Hannan (1982).

The motivation of this work is essentially to understand the effects of causality or noncausality in the estimation of b and how they affect the possibility of efficient estimation. This is why we choose to work with rather smooth models (with differentiability conditions on the distribution of the noise) in order to underline the role of causality in the results and the proofs. As a consequence of this work, the choice of a scaling factor will appear to be related to the causality of the model.

The organization of the paper is as follows: In subsection 2.1 we state the assumptions and derive the likelihood of the observations, which will be used throughout the paper. In subsection 2.2 we study the parametric situation, that is, the case where the distribution F of the noise sequence is known. For this purpose, we first show the LAN property for the model, which holds even when the process is noncausal. We then show how locally asymptotically minimax (LAM) estimates can be found when F is known. These results strengthen those of Breidt, Davis, Lii and Rosenblatt (1991), since the idea is that maximum likelihood achieves efficiency, but the LAM property is stronger. These parametric results are rather classical. The most interesting (and new) results are stated in subsections 2.3 and 2.4. Our main theorem in a negative result, conjectured in Gassiat (1990): When the distribution F is unknown, and with the special (usual) parameterization defined by (1), adaptive efficient estimates do not exist in the noncausal situation. In fact, they do exist if and only if the process is causal. We then show an asymptotic minimax theorem for the estimation of the parameter b, which quantifies what is lost because of the noncausality. The variance is shown to be the hardest "density parameter" to estimate. We then discuss these results in terms of the parameterization of the autoregressive process, and an issue is a new scaling of the parameter for which adaptive efficient estimation might be possible. Indeed, *scaling and causality together* lead to the impossibility of efficient adaptive estimation, and only one special scaling is necessary to avoid the problem of causality. All proofs are given in Section 3.

2. Main results.

2.1. *Notation, preliminaries.* For any filter $a = (a_k)_{k \in Z}$, we note by (A) the associated series: $A(z) = \sum_{k=-\infty}^{+\infty} a_k z^k$. If $X = (X_t)$, $t \in Z$, and the a_k are real and summable, the filtered process $Y = R_A X$ is defined by [see Azencott and Dacunha-Castelle (1984) for general definitions and properties]

$$Y_t = \sum_{k=-\infty}^{+\infty} a_k X_{t-k},$$

for all t in Z.

Now, with B defined in Section 1, we have

$$B(z) = \prod_{i=1}^{p} (1 - r_i z) = C(z) D(z),$$

with

$$C(z) = \prod_{i:|r_i|<1} (1 - r_i z) = \sum_{k=0}^{p_1} c_k z^k,$$

$$D(z) = \prod_{i:|r_i|>1} (1 - r_i z) = \sum_{k=0}^{p_2} d_k z^k,$$

where p_1, p_2 are integers such that $p_1 + p_2 = p$, c_k, d_k are all real, and all these numbers depend on b.

The parameter space is

$$\Theta = \{b \in R^p : B(z) \neq 0 \text{ if } |z| = 1\}$$

and Θ is open. Notice that $c = (c_k)$, $d = (d_k)$ are continuous functions of b, and differentiable functions of b in Θ. Now, $1/C$ and $1/D$ have the following development:

$$\frac{1}{C}(z) = \sum_{k=0}^{+\infty} \gamma_k z^k, \tag{2}$$

$$\frac{1}{D}(z) = \sum_{k=p_2}^{+\infty} \delta_k z^{-k}. \tag{3}$$

If the processes $Z = (Z_t)$ and $W = (W_t)$ are defined by

$$Z = R_D X, \qquad W = R_C X,$$

we have, using (1)–(3) and usual calculus in linear filtering,

$$Z = R_{1/C} U, \qquad W = R_{1/D} U,$$

so that

$$Z_t = \sum_{k=0}^{+\infty} \gamma_k U_{t-k}, \tag{4}$$

$$W_t = \sum_{k=0}^{+\infty} \delta_{k+p_2} U_{t+k+p_2}. \tag{5}$$

Notice that for (s_k) defined by

$$\frac{1}{B}(z) = \sum_{k=-\infty}^{+\infty} s_k z^k, \tag{6}$$

we have

$$X_t = \sum_{k=-\infty}^{+\infty} s_k U_{t-k}. \tag{7}$$

In the following, though C, D, γ, δ and s depend on b, this dependence will be omitted in the notation in order to make the paper readable. We will need the following additional assumptions:

(A3) F possesses a Lebesgue density f.

(A4) For all b in Θ, $((R_C X)_1, \dots, (R_C X)_{p_2})$ possesses a Lebesgue density $h(b, \cdot)$, and for all b in Θ, $((R_D X)_1, \dots, (R_D X)_{p_1})$ possesses a Lebesgue density $g(b, \cdot)$.

(A5) g and h are continuous in probability.

(A6) f is twice differentiable,

$$E(f'^2/f^2)(U) < \infty, \qquad E(U^2) < \infty,$$

$$E\big(U^2(f'^2/f^2)(U)\big) < \infty \quad \text{and} \quad x^2 f' \to 0 \quad \text{for} \quad x \to \infty.$$

(A7) Define $\phi(x) = (f''/f - f'^2/f^2)(x)$. For any b in Θ_0, on a small neighborhood W around b, there exists a function ω continuous at 0 [$\omega(0) = 0$] and a function Φ such that for c in W:

$$\|\phi((c * X)_t) - \phi((b * X)_t\| \le \omega(c - b)\,\Phi(X_t, \dots, X_{t-p})$$

and

$$E\big[U^2 \Phi(X_t, \dots, X_{t-p})\big] < +\infty.$$

These assumptions are quite natural when dealing with local asymptotic normality properties of the model. Notice that in case f is a sub- or super-Gaussian density, that is, if $f(x) = n(\rho)\exp(-\rho|x|^m)$, where $n(\rho) = [\int \exp(-\rho|x|^m)\,dx]^{-1}$, all these assumptions hold.

When necessary, we shall introduce the parameterization $\rho \to f_\rho$, $\rho \in \Xi$, and we shall make use of the following assumptions:

(A8) $E(D_\rho^1 \log f_\rho(U)) = 0$ and $E[(1/f_\rho^2)(\partial f_\rho/\partial_\rho)^2] < +\infty$.

(A9) Define

$$\lambda_\rho(x) = \left(\frac{D_\rho^2 f}{f} - \frac{\left(D_\rho^1 f\right)^2}{f^2} \right)(x)$$

(D_ρ^p is the pth derivative operator with respect to the variable ρ). For any (b, ρ) in $\Theta \times \Xi$, on a small neighborhood O around (b, ρ), there exists a function δ continuous at 0 [$\delta(0) = 0$] and a function Λ such that for (c, r) in O:

$$\|\lambda_r((c * X)_t) - \lambda_\rho((b * X)_t)\| \le \delta((c, r) - (b, \rho))\,\Lambda(X_t, \dots, X_{t-p})$$

and

$$E\big[U^2 \Lambda(X_t, \dots, X_{t-p})\big] < +\infty.$$

Notice again that in case $f_\rho = n(\rho)\exp(-\rho|x|^{2m})$, $m > 1$ (super-Gaussian family), assumptions (A8) and (A9) hold.

2.2. *LAN property for noncausal autoregressive models.* The main result of this subsection is the following theorem.

THEOREM 1. *Let* $(h_n) \subset R^p$ *be a bounded sequence and* $b^n = b^0 + n^{-1/2}h_n$. *Under assumptions* (A1)–(A7), *we have*

$$\log\frac{L_n(b^n)}{L_n(b^0)} - h_n^T \Delta_n(b^0) + \frac{1}{2}h_n^T\Sigma(b^0, f)h_n \to 0 \tag{8}$$

in b^0*-probability, and*

$$\Delta_n(b^0) \Rightarrow_{b^0} \mathscr{N}(0, \Sigma(b^0, f)), \tag{9}$$

where $\Rightarrow$ *denotes weak convergence, and for*

$$(\Delta_n(b))_k = \frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\left[\frac{f'}{f}\left(X_t + \sum_{l=1}^{p} b_l X_{t-l}\right)X_{t-k} + s_{-k}\right], \qquad k = 1,\dots,p,$$

$$(\Sigma(b, f))_{k,l} = \sum_{h\neq 0} s_{h-k}s_{-h-l} + \left(\sum_{h\neq 0} s_{h-k}s_{h-l}\right)E(U^2)E\left(\frac{f'^2}{f^2}\right)$$
$$+ s_{-k}s_{-l}\left(E\left(U^2\frac{f'^2}{f^2}\right) - 1\right).$$

Our aim is not to find minimal assumptions on the smoothness of the log-likelihood function, but to understand the role of causality in the estimation of the transfer function. The assumptions could be refined a little.

Notice also that the LAN property still holds in the case where F is Gaussian (see the proofs in subsections 3.1 and 3.2), but with the restriction that $\Sigma(b, f)$ is positive definite only for b in

$$\Theta_2 = \Theta_c \cup \Theta_a \cup \Theta_i,$$

where Θ_c is the set of causal parameters, Θ_a the set of purely noncausal parameters and Θ_i is the open set of parameters b for which the associated polynomial B satisfies the following property: If z is a root of B, z^{-1} is not a root of B.

From Theorem 1 we can build sequences of estimates which are locally asymptotically minimax (LAM) as is defined in Fabian and Hannan (1982). This can be done in the following way: Let b_n be a sequence of estimators for which $n^{1/2}(b_n - b)$ converges in distribution to some centered Gaussian distribution. Such a sequence exists; see Gassiat (1990). Let $\overline{b_n}$ be the nearest point to b_n in $\{(1/n^{1/2})(i_1;\dots;i_p),\ i_j \cap Z,\ j = 1,\dots,p\}$. This is a discrete sequence of estimators. Define now

$$\widehat{\overline{b_n}} = \overline{b_n} + \frac{1}{\sqrt{n}}\Sigma^{-1}(\overline{b_n}, f)\,\Delta_n(\overline{b_n}).$$

Recall that a sequence of estimators $\widehat{\overline{b_n}}$ is said to be regular if

$$\sqrt{n}\left(\widehat{\overline{b_n}} - b_0\right) - \Sigma(b_0, f)^{-1}\Delta_n(b_0) = o_{P_{b_0}}(1).$$

We have the following theorem.

THEOREM 2. *Under the assumptions of Theorem* 1, $\widehat{b_n}$ *is regular, and is thus LAM.*

Notice again that the theorem requires a preliminary sequence of estimators, converging with speed $n^{1/2}$, which exists in the Gaussian case only when the variance is known. An alternative method is the maximum likelihood estimator, which, as proved in Breidt, Davis, Lii and Rosenblatt (1991), is asymptotically Gaussian with variance Σ^{-1}. The LAM property is somewhat stronger.

The proofs of Theorems 1 and 2 are given in Section 3.

These estimators depend on the distribution F of the underlying white noise. This is why, in the next section, we investigate the existence of LAM estimators when this distribution is unknown.

2.3. *Existence of adaptive LAM estimates in the noncausal case.* Let $\mathscr{A}$ be the family of subproblems for which the density of the distribution of U is f_ρ, such that for any ρ, f_ρ verifies assumptions (A1)–(A7) and where the parametrization $\rho \to f_\rho$ verifies assumptions (A8) and (A9).

The main result concerning adaptive estimation in noncausal situations is a negative one.

THEOREM 3. *Under the assumptions described above, with the parameter defined by* (1): *As soon as*

$$E\left[U\left(\frac{f'_\rho}{f_\rho^2}\frac{\partial f_\rho}{\partial \rho}\right)(U)\right] \neq 0$$

and b is not causal, there exists no locally asymptotically minimax $\mathscr{A}$-adaptive estimator at b in the sense of Fabian and Hannan (1982).

For the super-Gaussian family, $f_\rho = n(\rho)\exp(-\rho|x|^{2m})$, $m > 1$,

$$E\left[U\left(\frac{f'_\rho}{f_\rho^2}\frac{\partial f_\rho}{\partial \rho}\right)(U)\right] = 2m\rho\operatorname{Var}(U^{2m}) \neq 0$$

and the theorem holds.

Kreiss (1987a, b) gives a construction of adaptive LAM estimates when b is causal, so that, according to the theorem, causality is a necessary and sufficient condition for the existence of adaptive LAM estimates. Notice that, as will become clearer later, Theorem 3 depends on the scaling chosen for the parameter in (1) (see Theorem 5 and subsection 2.6 below).

2.4. *Adaptive asymptotic minimax bounds in the noncausal case.* In this section we derive an asymptotic minimax bound for the estimation of the parameter b following semiparametric ideas developed in Begun, Hall, Huang and Wellner (1983). For this purpose, we need additional notation. As in Kreiss

(1987b) we define the following local parameter $(h, \beta) \in H_n$:

$$H_n = R^p \times \left\{\beta \in L_2(R) : \int \beta\sqrt{f} = 0, \int \beta^2 < n\right\},$$

$$b_n = b + \frac{1}{\sqrt{n}} h, \qquad h \in R^p,$$

$$f_n = \left[\sqrt{1 - \frac{1}{n}\int \beta^2\, dx}\,\sqrt{f} + \frac{\beta}{\sqrt{n}}\right]^2.$$

When n increases, H_n tends to $H = R^p \times \{\beta \in L_2(R)/\int\beta\sqrt{f} = 0\}$.

Let l: $R^p \to R^+$ be a loss function for the parameter b which is lower semicontinuous and subconvex. Let $\mathscr{K}$ denote the family of compact subsets of H. Let E_n denote the expectation with respect to the distribution defined by (b_n, f_n).

THEOREM 4.

$$\sup_{K \in \mathscr{K}} \liminf_{n\to\infty} \inf_{\widehat{b_n}} \sup_{(h,\beta)\in K} E_n l\left(\sqrt{n}\left(\widehat{b_n} - b\right)\right) \geq El(Z_*),$$

where Z_* *is a centered Gaussian random variable with variance* $M(b, f)$ *given by*

$$M(b, f) = (T_m + \mathrm{Var}(U) I(f) T_c)^{-1},$$

where

$$I(f) = E\left(\frac{f'^2}{f^2}(U)\right), \quad (T_m)_{k,l} = \sum_{h\neq 0} s_{h-k}s_{-h-l}, \quad (T_c)_{k,l} = \sum_{h\neq 0} s_{h-k}s_{h-l}.$$

REMARKS. $M(b, f)$ is well defined if and only if the true distribution F is not Gaussian, otherwise $T_m + T_c$ is not invertible [for a Gaussian centered distribution $\mathrm{Var}(U)I(f) = 1$]. If the true b is causal, $M(b, f) = (1/\mathrm{Var}(U)I(f))T_c^{-1}$, which is the minimax bound given by the LAN property and attainable using adaptive estimators; see Kreiss (1987a, b). $M(b, f)$ is a lower asymptotic bound for the variance of adaptive estimators.

The proof of Theorem 4 is given in Section 3. It relies on the following local asymptotic normality of the semiparametric model: If L_n is the local log-likelihood ratio:

$$L_n = \Delta_n(h, \beta) - \tfrac{1}{2}\sigma^2(h, \beta) + o_P(1),$$

where

$$\Delta_n(h,\beta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{p} \sum_t \left(\frac{f'}{f}(U_t) X_{t-k} + s_{-k} \right) h_k + \frac{1}{\sqrt{n}} \sum_t \frac{2\beta}{\sqrt{f}}(U_t),$$

$$\sigma^2(h,\beta) = h^T \Sigma h + \int 4\beta^2\, dx + 2\left(\sum_{k=1}^{p} h_k s_{-k} \right)\left(\int u \frac{f'}{\sqrt{f}}(u) 2\beta(u)\, du \right),$$

and where $\Delta_n(h,\beta)$ converges in distribution to the centered Gaussian distribution with variance $\sigma^2(h,\beta)$. We may then deduce the convergence of experiments to a Gaussian one, apply the Hájek–Le Cam minimax theorem and calculate the lower bound in the same manner as Millar (1981) or Begun, Hall, Huang and Wellner (1983), by orthogonal projection on the orthogonal of the nuisance space.

To understand better how the density parameter occurs to enlarge the asymptotic bound, we now make a calculation using one-dimensional parametrization of the nonparametric component. Suppose $\mathscr{A}$ is a family of subproblems as in subsection 2.3 which satisfies all assumptions of Theorem 3. In the model parametrized by (b,ρ), the asymptotic minimax variance of a sequence of estimators of b is the $p \times p$ matrix M_ρ with

$$M_\rho = [\tilde{\Sigma}^{-1}]_{11},$$

where $[\tilde{\Sigma}^{-1}]_{11}$ designates the $p \times p$ matrix $((\tilde{\Sigma}^{-1})_{k,l})$, $k = 1,\dots,p$, $l = 1,\dots,p$.

THEOREM 5. *The supremum of* M_ρ *(in the usual sense of ordering positive definite matrices) over all families* $\mathscr{A}$ *verifying the assumptions of Theorem* 3 *is* $M(b,f)$ *given in Theorem* 4. *This supremum is attained by the subfamily* $f_\rho(u) = \rho f(\rho u)$ *for which* ρ *is the "scale" parameter.*

The theorem says that the hardest one-dimensional subproblem is the one that estimates simultaneously b and the variance of the noise sequence. In other words, the estimation of the variance is the "density" estimation which is the most correlated with the estimation of b, and which makes more important the difficulty of estimating b when the distribution of U is unknown. An issue is to choose a scaling that is as much as possible uncorrelated with the variance and the parameter estimators; see subsection 2.6.

2.5. *Construction of adaptive asymptotic minimax estimators.* To see that the bound given in the above subsection is sharp, we must be able to build an adaptive sequence of estimators which achieves this bound. Following the ideas of Park (1991) in the i.i.d. case, a good candidate could be

$$\widehat{b_n} = \overline{b_n} + \hat{M} \frac{1}{\sqrt{n}} \sum_{t=p}^{n} \hat{a}(t),$$

with

$$\alpha(t) = \frac{f'}{f}(U_t)X_{t-} + s_-\left[1 - U_t\frac{f'}{f}(U_t) + \sqrt{f}(U_t)\right]$$

and $\hat{\alpha}(t)$ is $\alpha(t)$ with f and f' being replaced by a "good" estimator, and U_t by the convolution of X with $\overline{b_n}$ at index t.

Additional assumptions and extra work to investigate carefully the asymptotics are necessary, in order to check if this estimator is asymptotically minimax (i.e., which achieves the lower bound of Theorem 4). It will be the object of further work.

2.6. *Discussion.* In case b is purely noncausal and with the normalization condition $b_p = 1$ (instead of $b_0 = 1$), it is easy to see that the necessary condition of Fabian and Hannan for the existence of adaptive LAM estimates holds. Indeed, the only change in the proof of Theorem 3 is that $\tilde{\Sigma}_{i,p+1}$ is now proportional to s_{-i+1} (corresponding to the change of parameter). Now, as b is purely noncausal, $s_{-k} = 0$ for $k = 0,\dots,p-1$. In this situation, efficient adaptive estimation should be possible. In fact, it is possible, since with the normalization $b_p = 1$, the model is the usual causal one with time reversed, and the statistical properties are then the same.

More generally, suppose the parameterization is as follows: $b = (b_0,\dots,b_p)$, B is the associated polynomial, C and D are the polynomials such that $B = CD$, C has roots only outside the unit circle D has roots only inside the unit circle, $c = (c_0,\dots,c_{p_1})$ are the coefficients of $z^0,\dots,z^{p_1}$ in C, $d = (d_0,\dots,d_{p_2})$ are the coefficients of $z^0,\dots,z^{p_2}$ in D. Choose now the normalization condition $c_0 = 1$, $d_{p_2} = 1$ (which induces a very special normalization condition on b). Then, if the parameter is now (c,d), it appears that, following the same lines as above, the necessary condition of Fabian and Hannan holds, so that efficient adaptive estimation of (c,d) should be possible. In some sense, this normalization condition corresponds to the choice of parameterizing in the orthogonal of the scaling nuisance as described in subsection 2.4. It would be interesting to obtain adaptive efficient estimates of (c,d) in this model. This will be developed in further work. But it would, of course, lead to no improvement in the original model by rescaling.

3. Proofs. We give a complete proof of the local asymptotic normality of the model, and similar arguments can be used to prove the same property for subfamilies, which will be used for Theorems 3 and 4, where the proof will be omitted.

First of all, let us derive the exact form of the likelihood of the observations. Let $M_n(b)$ be the following matrix:

$$\begin{pmatrix} M_{n,1}(b) \\ M_{n,2}(b) \\ M_{n,3}(b) \end{pmatrix},$$

where $M_{n,1}(b)$ is the $p_2 \times n$ matrix:

$$\begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & c_{p_1} & \cdot & \cdot & c_1 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & c_{p_1} & \cdot & \cdot & c_1 & 1 \end{pmatrix}.$$

$M_{n,2}(b)$ is the $p_1 \times n$ matrix:

$$\begin{pmatrix} d_{p_2} & \cdot & \cdot & d_1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & d_{p_2} & \cdot & \cdot & d_1 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{pmatrix}.$$

$M_{n,3}(b)$ is the $n - p \times n$ matrix:

$$\begin{pmatrix} b_p & b_{p-1} & \cdot & \cdot & \cdot & \cdot & b_1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & b_p & \cdot & \cdot & \cdot & \cdot & \cdot & b_1 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & b_p & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}.$$

Using equations (4) and (5), it is immediate that $(R_C X)_{n+1-p_2}, \dots, (R_C X)_n$, $(R_D X)_{p_2}, \dots, (R_D X)_p$ are independent of the U_t for t varying from $p + 1$ to n. From the change of variables through M_n, the likelihood $L_n(b)$ of the observation is easily obtained

$$\begin{aligned} L_n(b) = &\left(\prod_{t=p+1}^{n} f\left(X_t + \sum_{k=1}^{p} b_k X_{t-k} \right) \right) h(b, (R_C X)_{n+1-p_2}, \dots, (R_C X)_n) \\ &\times g(b, (R_D X)_{p_2}, \dots, (R_D X)_p) |J_n(b)|, \end{aligned} \tag{10}$$

with $J_n(b) = \det(M_n(b))$.

Remarks. If b is causal, that is, if $s_k = 0$ for negative k, $D = 1$, $C = B$ and $J_n(b) = 1$. If b is anticausal, that is, if $s_k = 0$ for positive k, $D = B$, $C = 1$ and $J_n(b) = |b_p|^{n-p}$.

3.1. *Proof of the asymptotic normality of* $\Delta_n(b)$. We first calculate the asymptotic variance of $\Delta_n(b)$.

Lemma 1. $\mathrm{Var}_b(\Delta_n(b))$ *converges to* $\Sigma(b, f)$ *as stated in Theorem* 1.

PROOF. Direct computations using the development (7) lead to

$$E_b(\Delta_n(b)) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left(E_b\left[\frac{f'}{f}(U_t) \sum_{i=-\infty}^{+\infty} s_l U_{t-k-l} \right] + s_{-k} \right) = 0,$$

$$\begin{aligned} &E_b(\Delta_n(b)_k\, \Delta_n(b)_t) \\ &\quad = \frac{1}{n} \sum_{t,u=p+1}^{n} E\left(\frac{f'}{f}(U_t) \frac{f'}{f}(U_u) X_{t-k} X_{u-l} \right) \\ &\qquad - \frac{(n-p)^2}{n} s_{-k} s_{-l} \\ &\quad = \left[\frac{(n-p)(n-p-1) - (n-p)^2}{n} \right] s_{-k} s_{-l} \\ &\qquad + \frac{n-p}{n} \left[E\left(U^2 \frac{f'^2}{f^2}(U) \right) s_{-k} s_{-l} \right. \\ &\qquad\qquad \left. + E(U^2) E\left(\frac{f'^2}{f^2}(U) \right) \sum_{i \neq 0} s_{-k-i} s_{-l-i} \right] \\ &\qquad + \frac{1}{n} \sum_{t \neq u,\, t,u=p+1}^{n} s_{t-u-k} s_{u-t-l}. \end{aligned}$$

Now

$$\frac{1}{n} \sum_{t \neq u,\, t,u=p+1}^{n} s_{t-u-k} s_{u-t-l} = \frac{1}{n} \sum_{h \neq 0,\, h=p+1-n}^{n-p-1} (n-p-|h|) s_{h-k} s_{-h-l}$$

and, using the exponential convergence to 0 of hs_h when h goes to ∞, it is easy to see that the series $\sum_{h=-\infty}^{h=+\infty} |h| s_{h-k} s_{-h-l}$ is finite, and it follows that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t \neq u,\, t,u=p+1}^{n} s_{t-u-k} s_{u-t-l} = \sum_{h \neq 0} s_{h-k} s_{-h-l}.$$

Letting n go to ∞ in the expression of $E_b(\Delta_n(b)_k\, \Delta_n(b)_t)$ gives the lemma. □

LEMMA 2. *$\Sigma(b, f)$ is positive definite if and only if F is not Gaussian or b is in Θ_2.*

PROOF. For any $c = (c_1, \dots, c_p)$,

$$c^T\Sigma(b,f)c = \sum_{h\neq 0}(c*s)_h(c*s)_{-h} + \sum_{h\neq 0}(c*s)_h^2 EU^2E\frac{f'^2}{f^2} + (c*s)_0^2\left[E\left(U^2\frac{f'^2}{f^2}\right) - 1\right]$$

$$= \frac{1}{2}\sum_{h\neq 0}\left[(c*s)_h + (c*s)_{-h}\right]^2 + \sum_{h\neq 0}(c*s)_h^2\left[EU^2E\frac{f'^2}{f^2} - 1\right] + (c*s)_0^2\left[E\left(U^2\frac{f'^2}{f^2}\right) - 1\right].$$

(i) If F is not Gaussian, $EU^2E(f'^2/f^2) > 1$ and $E(U^2f'^2/f^2) > 1$, so that $c^T\Sigma(b,f)c = 0$ if and only if $c*s = 0$, and $\Sigma(b,f)$ is positive definite.

(ii) If F is Gaussian, $EU^2E(f'^2/f^2) = 1$ and $E(U^2f'^2/f^2) = 3$, so that $c\Sigma(b,f)c = 0$ if and only if $(c*s)_h + (c*s)_{-h} = 0$ for all h. If b is causal (resp. anticausal), $(c*s)_h = 0$ for negative h (resp. for positive h), so that the condition becomes $c*s = 0$, which implies $c = 0$. If not, the condition may be written with the associated polynomials

$$\frac{C(z)}{B(z)} + \frac{C(z^{-1})}{B(z^{-1})} = 0,$$

for all z in the complex plane, which is equivalent to

$$C(z)B(z^{-1}) + C(z^{-1})B(z) = 0.$$

If b is in Θ_i, the last equation implies that all roots of B are roots of C, which implies $C = B$ since they have the same degree. But this is impossible since 0 is a root of C and not of B. If B has at least two roots which are inverse ones of each other, $b = b_0 * b_1$ where b_0 is in Θ_i and b_1 is such that the associated polynomial has all roots two by two inverse related. The same reasoning as before shows that if w is a root of B for which w^{-1} is not a root of B, w is a root of C. So that, dividing by $z - w$ and $z^{-1} - w$ and using inductive reasoning, we have

$$C_1(z)B_1(z^{-1}) + C_1(z^{-1})B_1(z) = 0,$$

where $C_1 = C/B_0$. If the degree of B_1 is $2q$, this equation is equivalent to

$$C_1(z) + z^{2q}C_1(z^{-1}) = 0.$$

As soon as $q \geq 2$, this equation possesses nonzero solutions. □

To prove the asymptotic normality of $\Delta_n(b)$, we shall use a theorem of McLeish (1975) which says: Let $\mathscr{B}_n$ be a nondecreasing sequence of sigmaalgebras, V_n a process and $S_n = \sum_{i=1}^{n} V_i$. Assume:

(i) There exists a positive sequence g_k converging to 0 such that, for $i > 0$ and $k \geq 0$,

$$\|E^{i-k}V_i\|_2 \leq g_k,$$
$$\|V_i - E^{i+k}V_i\|_2 \leq g_{k+1},$$

where E^i denotes the conditional expectation with respect to $\mathscr{B}_i$, and $\|\cdot\|_2$ denotes the square root of the variance.

(ii) There exists a positive nondecreasing sequence R_n such that

$$\sum 1/(nR_n) < \infty, \qquad R_n - R_{n-1} = O\left(\frac{R_n}{n}\right), \qquad g_n = o\left(\frac{1}{\sqrt{n}\,R_n}\right).$$

(iii) The V_n are uniformly integrable.
(iv) $\mathrm{Var}(s_n/\sqrt{n})$ converges to σ^2.
(v) $E^{k-m}[(s_{n+k} - s_k)^2/n]$ converges to σ^2 when m, k and n go to ∞.

Then $S_n/n^{1/2}$ converges in distribution to the Gaussian distribution $\mathscr{N}(0, \sigma^2)$. Choose for $\mathscr{B}_n$ the sigma-algebra generated by $\{U_t,\ t \leq n\}$, and for any $\alpha = (\alpha_k)_{k=1,\dots,p}$, define

$$V_n = \sum_{k=1}^{p} \alpha_k\left(\frac{f'}{f}(U_n)X_{n-k} + s_{-k}\right).$$

Let $g_i = (EU^2)E(f'^2/f^2)^{1/2}(\sum_{k=1}^{p}\|\alpha_k\|)(\sum_{l=-\infty}^{-i}\|s_l\|)$ and $R_n = n$. The series $(\sum_{i=-\infty}^{-n}|s_l|)$ decreases exponentially fast to 0, so that (ii) holds.

For positive k,

$$E^{i-k}V_i = \sum_{h=1}^{p} \alpha_h s_{-h},$$

$$V_i - E^{i+k}V_i = \sum_{h=1}^{p} \alpha_h\left[\frac{f'}{f}(U_n)\sum_{l=-\infty}^{-h-k-1} s_l U_{n-h-l} + s_{-h}\right],$$

so that (i) follows by direct computations. Using similar calculations as for Lemma 1, (iv) follows with $\sigma^2 = \alpha^T\Sigma(b, f)\alpha$, (v) is a consequence of the law of large numbers, and (iii) obviously holds. We may conclude that $\alpha^T\Delta_n(b)$ converges in distribution to the centered Gaussian distribution with variance $\alpha^T\Sigma(b, f)\alpha$ and the convergence of $\Delta_n(b)$ follows. □

3.2. *Proof of* (8). To prove the LAN property of the model, let us introduce additional notation. $q_n(b, b^0)$ will be the approximation of the log-likelihood ratio

$$q_n(b, b^0) = \sum_{t=p+1}^{n} \log\frac{f\left(X_t + \sum_{k=1}^{p} b_k X_{t-k}\right)}{f\left(X_t + \sum_{k=1}^{p} b_k^0 X_{t-k}\right)} + \log|J_n(b)| - \log|J_n(b^0)|.$$

We then have

$$\log \frac{L_n(b)}{L_n(b^0)} = q_n(b, b^0) + V_n(b, b^0),$$

with

$$V_n(b, b^0) = \log \frac{h(b, (R_C X)_{n+1-p_2}, \dots, (R_C X)_n)}{h(b^0, (R_{C^0} X)_{n+1-p_2}, \dots, (R_{C^0} X)_n)} + \log \frac{g(b, (R_D X)_{p_2}, \dots, (R_D X)_p)}{g(b, (R_{D^0} X)_{p_2}, \dots, (R_{D^0} X)_p)}.$$

LEMMA 3. *For all b in Θ, we have the following derivatives, for $k = 1, \dots, p$ and $l = 1, \dots, p$,*

$$\frac{\partial}{\partial b_k} \log|J_n(b)| = (n-p) s_{-k},$$

$$\frac{\partial}{\partial b_l} s_{-k} = -\sum_{h=-\infty}^{+\infty} s_{h-k} s_{-h-l}.$$

PROOF. Suppose for a moment that the expectations of the gradients of the three log-likelihoods h, g and L_n are 0 (which is a weak realizable assumption on a neighborhood of any point b in Θ). Using the expression of L_n given by (10), we obtain

$$E\left[\frac{\partial}{\partial b_k} \log|J_n(b)| + \sum_{t=p+1}^{n} \frac{f'}{f}\left(X_t + \sum_{l=1}^{p} b_l X_{t-l}\right) X_{t-k}\right] = 0,$$

which gives, using the development (7) and usual computations,

$$\frac{\partial}{\partial b_k} \log|J_n(b)| = (n-p) s_{-k}.$$

Now, $|J_n(b)|$ is deterministic and independent of the f, g and h, so that the result holds also with no assumptions on f, g and h.

Write now that s is the inverse of b (taking $b_0 = 1$): $\forall\ h \in N$, $\sum_{k=0}^{p} b_k s_{h-k} = 1_{\{0\}}(h)$, where $1_{\{0\}}(\cdot)$ is the indicator function of the set $\{0\}$. Then

$$\sum_{k=0}^{p} b_k \frac{\partial s_{h-k}}{\partial b_l} + s_{h-l} = 0.$$

Define $c_u = \partial s_u / \partial b_l$, $v_h = -s_{h-l}$. The previous equation may be written as

$$b * c = v,$$

which may be inverted in

$$c = v * s$$

or, equivalently,

$$c_{-k} = \sum_h v_h s_{-k-h},$$

which is the second result of the lemma.

Using assumption (A5), we get

$$\log \frac{L_n(b^n)}{L_n(b^0)} = q_n(b^n, b^0) + o_P(1).$$

Now, using Lemma 3, we get

$$\nabla q_n(b, b^0) = \sqrt{n}\, \Delta_n(b),$$

so that the expression for the Taylor expansion to the second order of q_n becomes

$$q_n(b^n, b^0) = h_n^T \Delta_n(b^0) + \frac{1}{2} h_n^T \left[\int_0^1 \frac{1}{n} D_2 q_n(b^0 + u(b^n - b^0))\, du \right] h_n.$$

Using the expression of Δ_n, we get

$$(D_2 q_n(b))_{k,l} = \sum_{t=p+1}^{n} \left(\frac{f''}{f} - \frac{f'^2}{f^2} \right) \left(\sum_{h=0}^{p} b_h X_{t-h} \right) X_{t-k} X_{t-l} + (n-p) \frac{\partial}{\partial b_l} s_{-k}.$$

Now, using assumption (A7) it is easy to see that

$$\left[\int_0^1 \frac{1}{n} D_2 q_n(b^0 + u(b^n - b^0))\, du \right] - \frac{1}{n} D_2 q_n(b^0)$$

converges to 0 in b^0-probability. To prove (8), it then remains to show that $(1/n) D_2 q_n(b^0)$ converges to $-\Sigma(b^0, f)$ when n goes to ∞. Using Lemma 3,

$$\frac{1}{n} (D_2 q_n(b^0))_{k,l} = \frac{1}{n} \sum_{t=p+1}^{n} \left(\frac{f''}{f} - \frac{f'^2}{f^2} \right) (U_t) X_{t-k} X_{t-l} - \frac{n-p}{n} \sum_h s_{h-k} s_{-h-l}.$$

Using the law of large numbers and expansion (7), the result follows. □

3.3. *Proof of Theorem* 2. Using the terminology of Fabian and Hannan, it is enough to prove that $\widehat{b_n}$ is regular [Theorem 3, page 467 of Fabian and Hannan (1982)], that is,

$$\sqrt{n}\left(\widehat{b_n} - b_0\right) - \Sigma(b_0, f)^{-1} \Delta_n(b_0) = o_{P_{b_0}}(1).$$

Now

$$\sqrt{n}\left(\widehat{b_n} - b_0\right) - \Sigma(b_0, f)^{-1} \Delta_n(b_0) = A_n + B_n,$$

where

$$A_n = \sqrt{n}\left(\overline{b_n} - b_0\right) + \Sigma(b, f)^{-1}\left(\Delta_n(\overline{b_n}) - \Delta_n(b_0)\right),$$

$$B_n = \left[\Sigma(\overline{b_n}, f)^{-1} - \Sigma(b, f)^{-1} \right] \Delta_n(\overline{b_n}).$$

Keep in mind that $\Sigma(c, f)$ is a continuous function of the roots of C as soon as c is in Θ, so that $\Sigma(\overline{b_n}, f)$ is a consistent estimator of $\Sigma(b, f)$. Now using the Taylor expansion

$$\Delta_n(\overline{b_n}) - \Delta_n(b_0) = (\overline{b_n} - b_0)^T \int_0^1 \frac{1}{\sqrt{n}} D_2 q_n(b_0 + u(\overline{b_n} - b_0))\, du$$

and the results of subsection 3.2,

$$\Delta_n(\overline{b_n}) - \Delta_n(b_0) = \sqrt{n}\,(\overline{b_n} - b_0)^T \left[-\Sigma(b_0, f) + o_{P_{b_0}}(1)\right],$$

it follows that A_n and B_n are $o(1)$ in P_{b_0} probability. □

3.4. *Proof of Theorem* 3. Using the submodel where the parameter is (b, ρ) and with a proof following the one developed for Theorem 1 (this is why it will be omitted), we see that the submodel is LAN with $\tilde{\Delta}_n$ given by

$$\tilde{\Delta}_n = \left(\Delta_n; \left[\tilde{\Delta}_n\right]_{p+1}\right),$$

$$\left[\tilde{\Delta}_n\right]_{p+1} = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \frac{1}{f_\rho} \frac{\partial f_\rho}{\partial \rho}\left(X_t + \sum_{k=1}^{p} b_k X_{t-k}\right)$$

and asymptotic variance $\tilde{\Sigma}$ given by

$$\tilde{\Sigma}_{i,j} = \Sigma(b, f_\rho)_{i,j}, \qquad 1 \le i \le p, 1 \le j \le p,$$

$$\tilde{\Sigma}_{p+1,k} = s_{-k} E\left[U\left(\frac{f_\rho'}{f_\rho^2} \frac{\partial f_\rho}{\partial \rho}\right)(U)\right],$$

$$\tilde{\Sigma}_{p+1,p+1} = E\left[\frac{1}{f_\rho^2} \frac{\partial f_\rho^2}{\partial \rho}(U)\right].$$

Now, applying condition 8, page 474 of Fabian and Hannan (1982), a necessary condition for the existence of locally asymptotically minimax $\mathscr{A}$-adaptive estimators is

$$\tilde{\Sigma}_{p+1,k} = 0, \qquad k = 1, \dots, p,$$

which reduces to

$$s_{-k} = 0, \qquad k = 1, \dots, p,$$

as soon as $E[U(f_\rho'/f_\rho^2)(\partial f_\rho/\partial \rho)(U)] \neq 0$.

Now, as s is the inverse of b, for any positive l:

$$s_{-l} + \sum_{k=1}^{p} b_k s_{-l-k} = 0.$$

Therefore, by induction, the necessary condition is equivalent to $s_{-k} = 0$ for all positive k, and b is causal. □

3.5. *Proof of Theorem* 4. We follow the proof of Theorem 3.2 of Begun, Hall, Huang and Wellner (1983), and our notation is similar to theirs.

First of all, their assumption $\mathscr{S}$ holds since $\{\beta \in L_2(R): \int \beta\sqrt{f} = 0\}$ is a subspace of $L_2(R)$. On $H = R^p \times \{\beta \in L_2(R): \int \beta\sqrt{f} = 0\}$, define the following scalar product:

$$\langle (h_1, \beta_1); (h_2, \beta_2)\rangle_H = h_1^T \Sigma h_2 + 4\langle \beta_1, \beta_2\rangle_{L^2(R)} + 2\langle h_1, s_-\rangle_{R^p} \left\langle \beta_2, \frac{uf'}{\sqrt{f}}(u)\right\rangle_{L^2(R)}$$

$$+ 2\langle h_2, s_-\rangle_{R^p} \left\langle \beta_1, \frac{uf'}{\sqrt{f}}(u)\right\rangle_{L^2(R)}.$$

It induces the following norm on H:

$$\|(h, \beta)\|_H^2 = h^T \Sigma h + \int 4\beta^2\, dx + 2\left(\sum_{k=1}^{p} h_k s_{-k}\right)\left(\int u \frac{f'}{\sqrt{f}}(u) 2\beta(u)\, du\right).$$

As for Theorem 1, we can prove that, if L_n is the local log-likelihood ratio,

$$L_n = \Delta_n(h, \beta) - \tfrac{1}{2}\|(h, \beta)\|_H^2 + o_P(1),$$

where

$$\Delta_n(h, \beta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{p} \sum_t \left(\frac{f'}{f}(U_t) X_{tk} + s_{-k}\right) h_k + \frac{1}{\sqrt{n}} \sum_t \frac{2\beta}{\sqrt{f}}(U_t)$$

and where $\Delta_n(h, \beta)$ converges in distribution to the centered Gaussian distribution with variance $\|(h, \beta)\|_H^2$. We may then deduce the convergence of experiments to a Gaussian one. Suppose for the moment that $p = 1$. By classical projection theorem, there exists a β^* in $\{\beta \in L_2(R): \int \beta\sqrt{f} = 0\}$ minimizing $\|(1, \beta)\|_H^2$. This β^* satisfies $\langle (1, \beta^*), (0, \beta)\rangle_H = 0$ for all β in $\{\beta \in L_2(R): \int \beta\sqrt{f} = 0\}$. Calculation gives

$$\beta^*(u) = -s_{-1}\left(u \frac{f'}{\sqrt{f}}(u) + \sqrt{f}(u)\right).$$

Define now

$$M_{1,1} = \Sigma_{1,1} - s_{-1}^2\left(E\left(U^2 \frac{f'^2}{f^2}(U)\right) - 1\right).$$

Define also the mapping $\tau: H \to R$ by

$$\tau(h, \beta) = \langle (h, \beta); (1, \beta^*)/M_{1,1}\rangle_H.$$

Obviously, $\tau(h, \beta) = h$ and the adjoint τ^* of τ is given by

$$\tau^*(h) = \frac{h}{M_{1,1}}(1, \beta^*).$$

We now follow exactly Begun, Hall, Huang and Wellner (1983). The restriction of τ to the one-dimensional subspace $H^* = \{h(1, \beta^*),\ h \in R\}$ is linear, bounded, has dense range in R and is one to one. Moreover, $\|\tau^*(h)\|_H^2 =$

$h^2/M_{1,1}$, so the image law P_0 of the unit normal on H^* is simply $N(0, 1/M_{1,1})$. The inequality of Theorem 4 for the case $p = 1$ follows then from Proposition 3.1 of Millar (1979).

The case $p > 1$ requires slight modifications. The argument proceeds as before by defining $(h, \beta_1, \dots, \beta_k)$ in $R^p \times \{\beta \in L_2(R)\colon \int \beta\sqrt{f} = 0\}^k$, and projecting in all directions. We then obtain similarly

$$\beta_j^*(u) = -s_{-j}\left(u\frac{f'}{\sqrt{f}}(u) + \sqrt{f}(u)\right), \qquad j = 1, \dots, k,$$

and

$$M_{i,j} = \Sigma_{i,j} - s_{-i}s_{-j}\left(E\left(U^2\frac{f'^2}{f^2}(U)\right) - 1\right)$$

and Theorem 4 is proved. □

3.6. *Proof of Theorem* 5. Let $\Sigma(b, \rho)$ be the asymptotic variance defined in subsection 3.4 for the LAN property of the model parameterized by (b, ρ). We have

$$M_\rho^{-1} = \Sigma(b,\rho)_{11} - \Sigma(b,\rho)_{12}\Sigma(b,\rho)_{22}^{-1}\Sigma(b,\rho)_{21},$$

where $\Sigma(b, \rho)_{11}$ is the $p \times p$ matrix on the left top of $\Sigma(b, \rho)$, $\Sigma(b, \rho)_{22}$ is the 1×1 matrix on the right bottom of $\Sigma(b, \rho)$, $\Sigma(b, \rho)_{12}$ is the $p \times 1$ matrix on the left bottom of $\Sigma(b, \rho)$ and $\Sigma(b, \rho)_{21}$ is the transpose of $\Sigma(b, \rho)_{12}$. For $k = 1, \dots, p$ and $l = 1, \dots, p$,

$$\begin{aligned}\left[M_\rho^{-1}\right]_{k,l} = \sum_{h\neq 0} s_{h-k}s_{-h-l} &+ \left(\sum_{h\neq 0} s_{h-k}s_{h-l}\right)E_\rho(U^2)E_\rho\left(\frac{f_\rho'^2}{f_\rho^2}(U)\right)\\ &+ s_{-k}s_{-l}\Bigg[E_\rho\left(U^2\frac{f_\rho'^2}{f_\rho^2}(U)\right) - 1\\ &\qquad - \left(E_\rho\left(U\frac{f_\rho'}{f_\rho^2}\frac{\partial f_\rho}{\partial\rho}(U)\right)\right)^2\left(E_\rho\left(\frac{1}{f_\rho^2}\frac{\partial f_\rho^2}{\partial\rho}(U)\right)\right)^{-1}\Bigg].\end{aligned}$$

For any $c \in R^p$,

$$\begin{aligned}c^T M_\rho^{-1}c = \sum_{h\neq 0}(c*s)_h(c*s)_{-h} &+ \left(\sum_{h\neq 0}(c*s)_h^2\right)E_\rho(U^2)E_\rho\left(\frac{f_\rho'^2}{f_\rho^2}(U)\right)\\ &+ (c*s)_0^2\Bigg[E_\rho\left(U^2\frac{f_\rho'^2}{f_\rho^2}(U)\right) - 1\\ &\qquad - \left(E_\rho\left(U\frac{f_\rho'}{f_\rho^2}\frac{\partial f_\rho}{\partial\rho}(U)\right)\right)^2\left(E_\rho\left(\frac{1}{f_\rho^2}\frac{\partial f_\rho^2}{\partial\rho}(U)\right)\right)^{-1}\Bigg].\end{aligned}$$

Notice now that

$$H(\rho) = \left[E_\rho\left(U^2 \frac{f_\rho'^2}{f_\rho^2}(U) \right) - 1 \right.$$

$$\left. - \left(E_\rho\left(U \frac{f_\rho'}{f_\rho^2} \frac{\partial f_\rho}{\partial \rho}(U) \right) \right)^2 \left(E_\rho\left(\frac{1}{f_\rho^2} \frac{\partial f_\rho^2}{\partial \rho}(U) \right) \right)^{-1} \right]$$

$$= \mathrm{Var}_\rho\left(U \frac{f_\rho'}{f_\rho} \right) - \frac{\mathrm{Cov}_\rho\left(U \frac{f_\rho'}{f_\rho}(U); \frac{1}{f_\rho} \frac{\partial f_\rho}{\partial \rho}(U) \right)^2}{\mathrm{Var}_\rho\left(\frac{1}{f_\rho} \frac{\partial f_\rho}{\partial \rho}(U) \right)},$$

which is always nonnegative, so that

$$M_\rho^{-1} \geq M(b, f_\rho)^{-1}$$

in the usual ordering of positive definite matrices.

Now, let f be any density satisfying assumptions (A1)–(A7). Define

$$f_\rho(u) = \rho f(\rho u),$$

where ρ is a scale parameter. We have $f_1 = f$, so that, for ρ in a small neighborhood around 1, f lies in the family (f_ρ), and also f_ρ satisfies assumptions (A1)–(A9) for ρ in a small neighborhood around 1. Now direct calculation gives

$$E_\rho\left(U \frac{f_\rho'}{f_\rho^2} \frac{\partial f_\rho}{\partial \rho}(U) \right) = \frac{1}{\rho}\left[E_1\left(U^2 \frac{f'^2}{f^2} \right) - 1 \right],$$

$$E_\rho\left(\frac{1}{f_\rho^2} \frac{\partial f_\rho^2}{\partial \rho}(U) \right) = \frac{1}{\rho^2}\left[E_1\left(U^2 \frac{f'^2}{f^2} \right) - 1 \right],$$

$$E_\rho\left(U^2 \frac{f_\rho'^2}{f_\rho^2}(U) \right) = E_1\left(U^2 \frac{f'^2}{f^2} \right)$$

and it follows that

$$H(\rho) = 0$$

and the theorem is proved. □

Acknowledgments. The author wants to thank A. Mokkadem for helpful discussions and a referee for providing the reference by Breidt, Davis, Lii and Rosenblatt and suggesting the question of a new normalization.

REFERENCES

AZENCOTT, R. and DACUNHA-CASTELLE, D. (1984). Séries d'observations irrégulières, modélisation et prévision. *Techniques stochastiques* (Masson, ed.). Paris.

BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BENVENISTE, A., GOURSAT, M. and RUGET, G. (1980). Robust identification of a non minimum phase system: blind adjustment of a linear equalizer in data communications. *IEEE Trans. Automat. Control* **25** 3.

BREIDT, F. J., DAVIS, R. A., LII, K.-S. and ROSENBLATT, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *J. Multivariate Anal.* **36** 175–198.

FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **9** 459–478.

GASSIAT, E. (1990). Estimation semi-paramétrique d'un modèle autorégressif stationnaire multiindice non nécessairement causal. *Ann. Inst. H. Poincaré Probab. Statist.* **26** 181–205.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

KREISS, J. P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

KREISS, J. P. (1987b). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 27–98.

LII, K. S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* **10** 1195–1208.

MCLEISH, D. L. (1975). Invariance principle for dependent variables. *Z. Wahrsch. Verw. Gebiete* **32** 165–178.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252.

MILLAR, P. W. (1981). The minimax principle in asymptotic statistical theory. *Ecole d'Été de Probabilités de Saint-Flour XI. Lectures Notes in Math.* **976**. Springer, New York.

PARK, B. U. (1991). Asymptotic minimax estimation in semiparametric models. *Probab. Theory Related Fields* **88** 107–120.

ROBINSON, E. A. (1984). Statistical pulse compression. *Proceedings of the IEEE* **72** 1276–1289.

TOUATI, A. (1989). Principes d'invariance avec limites non browniennes. Thèse d'état, Univ. Paris-Nord.

LABORATOIRE DE STATISTIQUE
UNIVERSITÉ PARIS-SUD
URA D0743
BAT. 425
91405 ORSAY-CÉDEX
FRANCE